Linked Data based applications for Learning Analytics Research: faceted searches, enriched contexts, graph browsing and dynamic graphic visualisation of data

Ricardo Alonso Maturana
gnoss.com
*Piqueras 31, 4th floor
E-26006 Logroño. La Rioja. Spain
+34 941248905
riam@gnoss.com

María Elena Alvarado
gnoss.com*
elenaalvarado@gnoss.com

María José Ibáñez
gnoss.com*
mariajoseibanez@gnoss.com

Susana López-Sola
gnoss.com*
susanalopez@gnoss.com

Lorena Ruiz Elósegui
gnoss.com*
lorenaruiz@gnoss.com

ABSTRACT

We present a case of exploitation of Linked Data about learning analytics research through innovative end-user applications built on GNOSS, a semantic and social software platform. It allows users to find and discover knowledge from two datasets, Learning Analytics Knowledge (LAK) and Educational Data Mining (EDM), and also reach some related external information thanks to the correlation with other datasets. We used four additional datasets, either to supplement information or to generate enriched contexts: Dbpedia, Geonames, DBLP-GNOSS (an index of scientific publications in Computer Science from DBLP) and DeustoTech Publications (publications of the Institute of Technology of the University of Deusto, and more specifically a selection of works by the DeustoTech Learning research unit). The featured applications are: faceted searches, enriched contexts, navigation through graphs and graphic visualization in charts or geographic maps. Faceted searches can be performed on three basic items: scientific publications, researchers (authors of the publications) and organizations in the learning analytics area. The search engine enables aggregated searches by different facets and summarization of results for each successive search. Analytics on data are provided firstly through that summarization given for results in every facet, and secondly through dynamic graphic representations for some attributes. Several charts are available to show the distribution of publications depending on different attributes (e.g. per publication type and year or per organization). The search results for organizations and researchers can be visualized in geographic maps.

This work was presented to the LAK Data Challenge 2013.

Categories and Subject Descriptors

Information systems - Data management systems - *Database design and models - Graph-based database models*

Information systems - Information systems applications - *Collaborative and social computing systems and tools - Social networking sites*

Classification Scheme: The 2012 ACM Computing Classification System (CCS) http://dl.acm.org/ccs.cfm

General Terms

Algorithms, Documentation, Standardization, Languages.

Keywords

Learning Analytics, Educational Data Mining, Semantic Web, Linked Data, linked open data, faceted search, semantic contexts, recommendation systems, geographic visualisation, geolocated data, discovery information systems, knowledge management, semantic data platform, end-user semantic platform, GNOSS.

1. MOTIVATION: PURPOSES OF THE SOLUTION

The main purpose of the service developed on the gnoss.com[1] software platform is to provide end-users with innovative applications that allow them to find and discover knowledge related to learning analytics research from the LAK and EDM datasets [1].[2] Based on the exploitation of Linked Data [2, 3], the system includes faceted searches, recommendation systems and adapted contexts. More specifically, the software solution serves the following purposes:

1. Explore and navigate the datasets (LAK and EDM) through faceted searches and graph browsing. It enables to find publications, researchers and organizations in the area, as well as to know about which topics are being investigated, who is working in which fields, where those people and their organizations are located, who has published about LAK in an organization, or who is collaborating with whom, for instance.

2. Access a geographic visualisation of researchers and organizations working in learning analytics, with the option of filtering results by different and aggregated facets.

3. Visualise dynamic charts of some analytic information related to the evolution and distribution of publications. The charts are dynamic as the results evolve in the chart with the successive selected filters in the search facets.

4. Discover related information within the dataset once the user has access a specific item (internal context), such as related LAK publications, co-authors, related and nearest organizations, etc.

[1] GNOSS: http://www.gnoss.com/en/about-gnoss

[2] LAK and EDM datasets are available online in: http://www.solaresearch.org/resources/lak-dataset

5. Discover external related information through the correlation with other datasets. Some examples of datasets have been chosen for the demo to show the potential of the GNOSS platform tools.

6. Facilitate the potential relevant re-use of these datasets as contexts in other scenarios by linking them to existing social learning communities on gnoss.com, or any community related to the study of those topics.

2. DESCRIPTION OF THE SOLUTION AND DATASETS

The solution exploiting the LAK datasets has been developed on gnoss.com, a social and semantic platform with a deep focus on the generation of social knowledge ecosystems and end-user applications in a Linked Data environment. It includes faceted searches, recommendation systems and adapted contexts in education, university and enterprises. GNOSS could be conceived as a network of networks or a linked networks space oriented to using semantic technologies for data and service integration. It expresses the data generated by users with default basic semantic standard vocabularies. This is done automatically when a user shares content on the platform. Besides, GNOSS has an engine for developing specific ontologies and, as a consequence, specific search engines if necessary. Moreover, it has a wide range of configurable social tools, which have been mostly deactivated for this demo, except for comments and the option to share the link via email or other social networks.

2.1 The basis: LAK and EDM datasets

The baseline information to develop the solution was obtained from two datasets related to learning analytics: 1) Learning Analytics and Knowledge (LAK) 2011-2012 and 2) Educational Data Mining (EDM) 2008-2012. Both of them have information about people (researchers), organizations in which they work, and publications (proceedings, inproceedings and articles).

The information of the original datasets was enriched with data coming from Dbpedia[3] and Geonames,[4] and also with automatically generated tags. Moreover, some duplicated data (researchers and organizations) that appeared when unifying the two datasets were eliminated.

This information was uploaded to an online space inside the gnoss.com platform to consume and exploit the data and present the end-user applications.

We prepared a general navigation through tabs that includes a homepage with content selection and three other tabs corresponding to the three entities from the datasets: publications, researchers and organizations.

The three previous entities were represented on the platform with their specific ontologies thanks to the semantic CMS of GNOSS following the standard vocabularies of the original data: FOAF (Friend-of-a-Friend),[5] SWRC (Semantic Web for Research Communities),[6] DC (Dublin Core),[7] etc. In addition, other vocabularies were included for representing the extended information and/or correlating datasets, for instance SIOC (Semantically Interlinked Online Communities),[8] SKOS (Simple Knowledge Organization System),[9] DBPROP[10] or GN (Geonames).[11]

2.2 Other datasets

Besides the direct consumption of the information provided in the datasets on learning analytics research, we used other additional datasets, either to supplement the information of the former ones (as explained above) or to automatically generate dynamic contexts with external related information. The following additional datasets were employed:

1. Dbpedia, for supplementing the data of organizations and obtaining geographic information that enables the connection with Geonames.

The automation of this process gave rise to incomplete information for some items, in such a way that we could not obtain the necessary information to represent all the researchers and organizations in the geographic map. As a consequence, the presentation of results differs from the 'mosaic view' (it includes all the results) and the 'map view' (it only represents the geolocated data). These data could be improved in the future.

This is a common problem in the Web of Data. Datasets usually need to be refined because of one or more of the following reasons: incomplete data, insufficient (missing) data or inconsistent data (data are not well described or depicted or are named wrongly). This complicates to provide an adequate service and, specially, this makes it difficult to upload datasets and set relations between data.

2. Geonames, with the aim of recovering geolocation data and use them to develop the exploitation of geographic visualization of data.

3. Two GNOSS existing datasets of scientific publications that we found interesting as contexts in the field of learning analytics: DBLP-GNOSS[12] and DeustoTech publications. [13]

'DBLP-GNOSS' is and index with over two million scientific publications in IT, developed by GNOSS in collaboration with the University of Deusto. The data of DBLP-GNOSS have been obtained from the dataset in the LOD cloud 'DBLP' promoted by the University of Trier, and have been enriched with abstracts and key words.

'DeustoTech publications' is the dataset of scientific publications of the Technology Institute of the University of Deusto, DeustoTech. As a demo of a relevant external context, we included a selection of the publications produced by the research team DeustoTech Learning.

[3] Dbpedia: http://dbpedia.org/

[4] Geonames: http://www.geonames.org/

[5] FOAF Vocabulary specification: http://xmlns.com/foaf/spec/

[6] SWRC ontology: http://swrc.ontoware.org/ontology#

[7] Dublin Core terms: http://purl.org/dc/terms/

[8] SIOC Core Ontology Specification: http://rdfs.org/sioc/spec/

[9] SKOS namespace: http://www.w3.org/2004/02/skos/core#

[10] Dbpedia ontology: http://dbpedia.org/Ontology

[11] Geonames ontology: http://www.geonames.org/ontology

[12] GNOSS Research Groups: http://researchgroups.gnoss.com

[13] DeustoTech publications: http://deusto.gnoss.com/comunidad/DeustoTech/Publications

2.3 Faceted searches

The web of structured data makes it possible to develop strategies for intelligent information retrieval based on faceted searches [4, 5, 6 and 7]. GNOSS has a powerful faceted search engine that is generated by the GNOSS semantic graphs (RDF triplets); the search engine exploits that graphs through reasoned or inference-based searches.

The main advantages of facet-based searches are:

- They are based on meaning and concepts, and relations between them.
- Users obtain reduced lists of results based on semantic properties or attributes of the data.
- They allow reasoning: a new search allows restricting the subset of data from the previous search across multiple facets. You can progressively filter results until you reach a manageable data set.

Searches on the LAK Data Challenge space in GNOSS can be started from two approaches:

- As a meta-search, seeking in any kind of content.
- Or selecting the item type to perform the search, either choosing it from the facet 'item type' in the home webpage, or navigating through the corresponding tab for every item. In this case, there are three basic item types: publications, researchers and organizations.

Once an item type is selected, the search engine provides specific facets for each of them, which are configurable in function of the available data. The relevant facets that have been set for each case of the LAK Data are:

- For publications: categories, tags, author, year, conference and publication type.
- For researchers: categories, tags, affiliation (organization) and country.
- For organizations: categories, tags, country, region, city and students number.

The GNOSS faceted search engine allows concatenated searches, and all relationships among the facets are recalculated with each successive filter for the corresponding set of results.

2.3.1 Summarization of results: direct quantitative exploitation of data

GNOSS offers summarization of the number of results in each property represented in the facets. The values are recalculated for every set of results in aggregated searches. This gives direct analytic information that is represented in the form of facets for searches (see example in Figure 1).

Thus, the search results give a lot of information through the facets: how they relate to the other searching attributes. For example, you look for publications with the tag 'intelligent tutoring system', you obtain 12 results and know who worked on this topic and who published the most papers, and you know that the author Zachary A. Pardos, for example, wrote 3 publications in the field. If you select this author, all the facets are recalculated and you can see how they relate to the publications about intelligent tutoring system by Pardos, for example, that they were published in 2008, 2011 and 2012, that one of them is related to Bayesian knowledge, and that he collaborated with other four authors.

Figure 1 Example of facets showing summarization of results

2.4 Navigation trough graphs and relationships between entities and properties

The possibilities of navigation through graphs that connect entities and properties (among them and with each other) are immense and n-dimensional. Just to give some examples of items relationships and possible navigation paths in the present case:

- Authors and papers: authors who wrote articles on a specific topic, authors of a publication you are interested in, papers written by a selected author.
- Researchers and organizations in which they work: related organizations working on similar topics.
- Authors and co-authors: if you find a researcher, you could be interested in the people working with him, and then discover other research areas the latter are working at, and see their location in a map.
- Related topics and their relation with researchers (authors): you look for a key word and you see other related terms and the researchers publishing on that subject. You can navigate through the authors and discover new publications, co-authors, etc.
- Location, people and research topics: you look for researchers by geographic criteria, e.g. United Kingdom, and you get the topics they are working at (tags).

2.5 Enriched contexts of information and recommendations

The Web of Data also enables to connect information significantly, which can be exploited in GNOSS for the generation of dynamic contexts that can be customized for each case.

In the present work on LAK and EDM data, we set several demonstration contexts depending on the object or entity that the user is viewing:

1. Contexts for the entity 'publication': related LAK and EDM publications (internal), related DBLP publications (external), DeustoTech Learning publications (external).

2. Contexts for the entity 'researcher': co-authors (internal), related organizations by topics (internal) and related DBLP publications (external).

3. Contexts for the entity 'organization': related organizations (internal), related researchers (internal), nearest organizations (internal), geolocation (external) and its visualization on a map.

4. Contexts of general purpose: Freebase definitions of tags of the contents (when the concepts have an article in Freebase). For example, if you select a publication about data mining, when you put the mouse on the tag 'data mining', a window appears with its definition on Freebase and the link to the Freebase and Wikipedia articles.

2.6 Geographical visualisation of data

The present work includes the development of an application to represent a set of geolocated results in a geographical map. In the case of LAK and EDM datasets, this visualisation is enabled for researchers and organizations (see example in Figure 2), combined with the option of filtering results by different and aggregated facets.

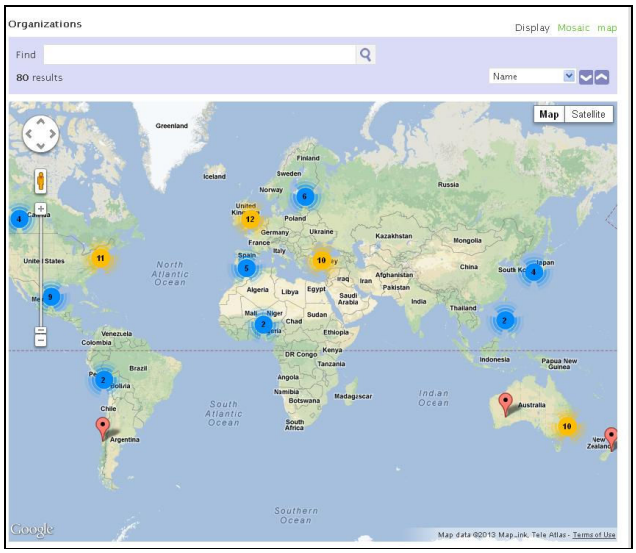

Figure 2 Geographic visualisation of search results (organizations)

2.7 Visualisation of analytics with dynamic charts

The analytics provided by summarization in search facets was supplemented with some graphic visualisations. Google charts tools[14] were integrated in the platform to represent some analytics related to the evolution and distribution of publications. Four types of charts were used: column chart, intensity map, pie chart and bar chart. The user can choose among several charts, and continue filtering through facets successively, thus seeing how the results evolve in the chart with the selected filters. Six charts were included to analyse LAK data:

- Evolution of number of publications per year and publication type (column chart, Figure 3). It shows how the number of publications in this area has increased during the last years, and how they are distributed in inproceedings (the main part), articles and proceedings. These results can be restricted to selected criteria filtering through facets.

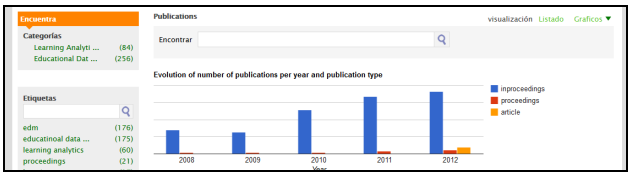

Figure 3 Evolution of number of publications per year and publication type

- Evolution of number of inproceedings per conference (column chart).
- Distribution of number of publications per country (intensity map, Figure 4). It gives a quick idea about the countries with more scientific production in the field, according to search criteria (total number, one or more specific topics, a selected year, etc.).

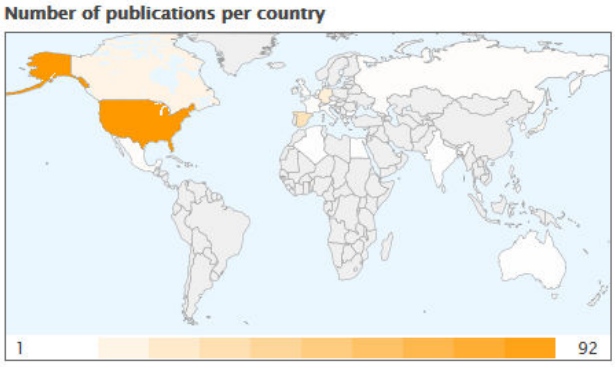

Figure 4 Distribution of number of publications per country (intensity map)

- Distribution of number of researchers with publications per year (pie chart, Figure 5).

[14] Google chart tools. Information for developers available in https://developers.google.com/chart/.

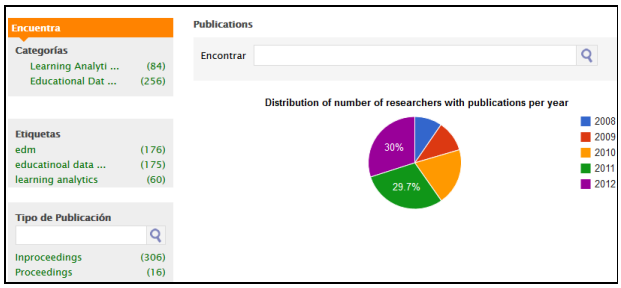

Figure 5 Distribution of number of researchers with publications per year (pie chart)

- Distribution of number of publications per organization (bar chart, Figure 6). The first view shows the total number of publications for each organization along the years, and shows clearly which ones have produced the larger amount, with the Worcester Polytechnic Institute leading the list. By filtering through facets, like year, tags or publication type, the user can observe how the chart changes depending on those filter options.

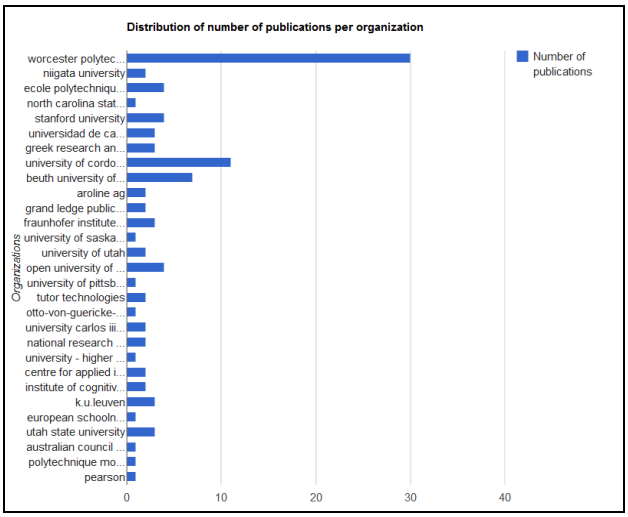

Figure 6 Distribution of number of publications per organization

- Distribution of number of publications per author (bar chart). It is similar to the previous one, but representing authors instead of organizations.

This work shows some examples of charts representing analytics on the LAK data, and it is extensible to additional similar exploitations.

3. LINK TO THE PLATFORM

The demo GNOSS solution for Learning Analytics Research is available in the following link: http://datasetexplorer.gnoss.com/en/community/LAKChallenge

4. ACKNOWLEDGMENTS

Government of Spain, Ministry of Economy and Competitiveness, CDTI (Centre for Industrial Technological Development (CDTI), for the funding of a project financed by CDTI (Project IDI-20110600).

University of Deusto, for the collaborative work in the generation of external contexts.

5. REFERENCES

[1] Taibi, D., Dietze, S., Fostering analytics on learning analytics research: the LAK dataset, Technical Report, 03/2013, URL: http://resources.linkededucation.org/2013/03/lak-dataset-taibi.pdf.

[2] Bizer, C.; Heath, T.; Berners-Lee, T. 2009. Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*, 5(3), 1-22. DOI: 10.4018/jswis.2009081901.

[3] Bauer, F.; Kaltenböck, M. 2012. *Linked Open Data: The Essentials. A Quick Start Guide for Decision Makers.* Edition Mono/monochrom, Vienna, Austria. ISBN: 978-3-902796-05-09.

[4] Suominen, O.; Viljanen, K.; Hyvönen, E. 2007. User-centric faceted search for semantic portals. In *The Semantic Web: Research and Applications. Proceedings of the 4th European Semantic Web Conference ESWC2007, forth-coming* (Innsbruck, Austria, June 3-7, 2007), 356-370. DOI: 10.1007/978-3-540-72667-8_26.

[5] Stefaner, M: Ferré, S.; Perugini, S.; Koren, J.; Zhang, Y. 2009. User Interface Design. In *Dynamic Taxonomies and Faceted Search*. 2009. Sacco, G. M.; Tzitzikas, Y. (Eds.)

[6] Ferré, S.; Hermann, A.; Ducassé, M. 2011. Semantic Faceted Search: Safe and Expressive Navigation in RDF Graphs. Research report. ISSN: 2102-6327.

[7] Dal Mas, M. 2012. Faceted Semantic Search for Personalized Social Search. In *Computing Research Repository*, 2012, abs-1202-6685. URL: http://arxiv.org/abs/1202.6685.